DEBREBERHAN UNIVERSITY

COLLEGE OF NATURAL AND COMPUTATIONAL SCINCES

POST GRADUATE STUDIES

DEPARTMENT OF STATISTICS

SURVIVAL ANALYSIS OF TIME TO RECOVERY AND ASSOCIATED FACTORS OF TUBERCULOSIS PATIENTS IN DESSIE COMPREHENSIVE SPECIALISED HOSPITAL, SOUTH WOLLO, ETHIOPIA.

By:

HAWULET MEKONNEN

A THESIS SUBMITTED TO THE SCHOOL OF POST GRADUATE STUDIES, COLLEGE OF NATURAL AND COMPUTATIONAL SCIENCES IN THE PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE MASTER OF SCIENCE DEGREE IN BIOSTATISTICS

JUNE, 2021
DEBRE BERHAN

ETHIOPIA

DEBREBERHAN UNIVERSITY

COLLEGE OF NATURAL AND COMPUTATIONAL SCINCES

POST GRADUATE STUDIES

DEPARTMENT OF STATISTICS

SURVIVAL ANALYSIS OF TIME TO RECOVERY AND ASSOCIATED FACTORS OF TUBERCULOSIS PATIENTS IN DESSIE COMPREHENSIVE SPECIALISED HOSPITAL, SOUTH WOLLO, ETHIOPIA.

By:

HAWULET MEKONENN

ADVISOR: BEZAREDE M. (ASSI.PROF.)

CO-ADVISOR: SINTAYEHU W.(MSc.)

A THESIS SUBMITTED TO THE SCHOOL OF GRADUATE STUDIES, COLLEGE OF NATURAL AND COMPUTATIONAL SCIENCES IN THE PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE MASTER OF SCIENCE DEGREE IN BIOSTATISTICS.

JUNE, 2021

DEBREBERHAN

ETHIOPIA

Declaration

I declare that the research work presented in this thesis entitled "Survival Analysis of time to recovery and Associated factor of Tuberculosis patients in Dessie Comprehensive Specialised Hospital" has been carried out by me and this work, or part thereof, has not been submitted for the award of any Degree or Diploma of this or any other university.

Name: Hawulet Mekonnen Signature: _____

Date of Submission: _____

This thesis has been submitted for examination with my/our approval as university advisor(s).

Advisor	Signature	Date
_____	_____	_____
Co-Advisor	Signature	Date
_____	_____	_____

APPROVAL SHEET

DEBRE BERHAN UNIVERSITY

COLLEGE OF NATURAL AND COMPUTATIONAL SCIENCES

SCHOOL OF GRADUATE STUDIES

DEPARTMENT OF STATISTICS

SURVIVAL ANALYSIS OF TIME TO RECOVERY AND ASSOCIATED FACTORS OF TUBERCULOSIS PATIENTS IN DESSIE COMPREHENSIVE SPECIALISED HOSPITAL, SOUTH WOLLO, ETHIOPIA.

Approved by the board of examiners

_____ _____ _____

Department chairman **signature** **date**

_____ _____ _____

Advisor **signature** **date**

_____ _____ _____

Co-advisor **signature** **date**

_____ _____ _____

Internal Examiner **signature** **date**

_____ _____ _____

External Examiner **signature** **date**

ACKNOWLEDGMENTS

First and foremost, I would like to acknowledge the one and the true God" Allah" for blessing me with the ability and guidance to succeed my MSc study over the past years. This study would never have been completed with out the contribution of many people to whom I would like to express my thinks. My special gratitude goes to my main advisor: Bezarede Mekonnen (ASSI.PROF.) and Co- Advisor:Sentayehu Workineh(MSc.) for their kind support, assistance, valuable comments and advice through the initial to final level, which enable me to develop understanding of the subject. I also extend my sincere gratitude to my instructors, staffs of Statistics Department and friends for their unreserved knowledge sharing and cooperation. My gratitude also goes to Samara University which sponsored my study. Next to this my heartfelt gratitude goes to my Husband, Mr. Kalid Abegaz for his motivation and support throughout the study. Finally, I am highly grateful to all staff members of Dessie Comprehensive Specialized Hospital for their permission and cooperation to provide all relevant information.

TABLE OF CONTENTS

CONTENT **page**

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AFB	Acid-Fast Bacilli
AFT	Accelerated Failure Time
AIC	Akaike Information Criteria
ANC	Antenatal Care
DOTS	Directly Observed Treatment Short course
DCSH	Dessie comprehensive specialised hospital
EPTB	Extra-pulmonary Tuberculosis
FMOH	Federal Ministry Of Health
LNTBL	lymph Node Tuberculosis
LTB	Latent Tuberculosis
LTBI	Latent Tuberculosis Infection
MDRTB	Multi Drug Resistant Tuberculosis
MTB	Mycobacterium Tuberculosis
NTBP	National Tuberculosis Programs
NTCP	National Tuberculosis Control Programme
PTB	Pulmonary Tuberculosis
TB	Tuberculosis
WHO	World health organization

ABSTRACT

Tuberculosis is one of the leading causes of mortality worldwide. The recovery time for tuberculosis patients taking different treatments varies among treatment types due to any other factors. The main objective of this study was to identify determinant factors of time to recovery from tuberculosis patients at Dessie comprehensive specialized hospital, south wollo, Ethiopia. It is a hospital based retrospective cohort study among tuberculosis patients under follow up from January 2017 to December 2020. A total of 311 Tuberculosis patients were included in the study. Secondary data was extracted from patients monitoring card and register book. Survival analysis to assess the stated objectives of this study. *Non parametric were used to compare the survival experience of different categories, semi-parametric survival model and AFT models were employed to identify time to recovery of the TB patients. From 311 patients, 130(41.80%) recovered in the follow-up period. The median time to recovery from TB was 270 days which was approximately nine months. Based on AIC, the log logistic accelerated failure time model is better to fit the data than other AFT model. The log-logistics AFT model result shows that the time to recovery of TB patients significantly affected by Age, work area (health facility and non-health facility), disease classification, initial weight, type of TB, WHO clinical TB stage (stage I and II) and HIV status. Covariates like work area, WHO clinical stage, disease classifications and HIV were shorter time to recover and the other covariates Type of TB, age and weight were prolonging time to recovery. I recommended as The ministry of health and policymakers should work on awareness about the risk factor of TB.*

Keywords*: Tuberculosis, log logistic Accelerated failure time, recovery time*

1. INTRODUCTION

1.1.Background of the study

Tuberculosis is chronic infectious disease caused by bacterium, Mycobacterium tuberculosis and Mycobacterium bovines which affects any parts of the body, mostly affects lung (Thacker et al., 2012).It is transmitted from patients with active tuberculosis through airborne droplet and from cattle infected with tuberculosis during raw milk consumption. It can affect any part of the body by which it is categorized as pulmonary and extra-pulmonary tuberculosis based on the organ it involves.

Pulmonary tuberculosis (PTB) is an infectious disease of the lung with airborne transmission that is associated with high morbidity and mortality worldwide and Extra-pulmonary (EPTB): It is the result of the spread of tuberculosis to other organs, most commonly pleura, lymph nodes, spine, joints, genitor-urinary tract, nervous system or abdomen (Silva et al., 2010). It is a disease with well-known risks factors, clinical features, diagnostic techniques and treatment modalities. The diagnostic technique are sputum smear microscopy developed 10 decades ago which detect bacteria from sputum sample, rapid molecular test to diagnose Tuberculosis (TB) and drug resistance TB. Countries with developed laboratory capacity diagnose TB using culture technique which is the current gold standard (WHO, 2015).

Tuberculosis mostly affects adults in their most productive years. However, all age groups are at risk. Over 95% of cases and deaths are in developing countries(WHO, 2020), Tuberculosis attacks the lungs but can affect other parts of the body. It is an air born disease which spread through coughing, sneezing and spitting by infected persons (Konstantin, 2010).

Tuberculosis (TB) disease remains a major public health problem which affects all age groups globally (Reves and Angelo, 2016). It is one of the top 10 causes of death and the leading cause from a single infectious agent and millions of people continue to fall sick with TB each year (WHO, 2018). In developing countries where population is dense and hygienic standards are poor, tuberculosis remains a major fatal disease. The high frequency of Mycobacterium tuberculosis in Sub-Sahara African countries as a result of poor nutrition, inadequate TB control measures, inadequate program for the disease and rapid growth of the population (Addo et al., 2010) .

Ethiopia is among the top 20 high TB burden countries (WHO,2017), The treatment success rate of TB among 30 high TB burden countries varied from 71% in Congo to 94% in Cambodia, with an 80% success rate in Ethiopia (WHO,2017), Overall, TB is the third cause of hospital admissions and the second cause of death in Ethiopia (Health, 2013).

A study conducted in Gondar, Ethiopia showed that the treatment success rate of tuberculosis patients was unsatisfactorily low and it became a serious public health concern ((Health, 2012). Being male, older age (Shen et al., 2009), smear-negative pulmonary or extra pulmonary tuberculosis and being rural resident was associated with lower treatment success rate ((Health, 2012) Therefore, TB control programs should be strengthened to prevent unnecessary deaths (Shen et al., 2009).Despite significant improvements in the TB treatment outcomes following the introduction of DOTS services in Ethiopia, many regions continue to report a large number of unsuccessful treatment outcomes: in the Amhara region 39.9% (Biruk et al., 2016), in the Tigray region 10.8% (Berhe et al., 2012), in the South region 14.8% (Gebrezgabiher et al., 2016) and in Addis Ababa 9.02% (Getahun et al., 2013). Therefore, there exists a critical need to identify the causes and predictors of unsuccessful TB treatment outcomes in Ethiopia.

According to Amhara regional state health office (ARSHO) report the number of TB cases is high (280 per 100,000) while case detection is very low (29%)(Amhara Regional state Health office) (ARSHO,2014).

In Amhara Region DOTS strategy has been introduced since 1996 and its geographic coverage reached 100% of the region in 2009 (Gebreegziabher et al., 2016). Study conducted in India has shown clearly that the prevalence of pulmonary TB disease is significantly higher among males than females (Muvunyi et al., 2010).

Globally, in 2016 there were an estimated 1.3 million TB deaths among HIV negative people and an additional 374,000 deaths among HIV-positive people. An estimated 10.4 million people (90% adults, 65% male, 10% people living with HIV) fell ill with TB in 2016. Regionally, the fastest declines in the TB mortality rate are in the WHO European Region and the WHO Western Pacific Region (six and four point six percent per year, respectively, since 2010). High TB burden countries with rates of decline exceeding six percent per year since 2010 include Ethiopia, the Russian Federation, the United Republic of Tanzania, Viet Nam and Zimbabwe and 98% of TB

deaths are in the developing world affecting mostly young adults in their most productive years most of them are in Africa (WHO, 2018, Terefe and Gebre wold, 2018). This will be studied by means of real dataset which is collected from tuberculosis (TB) patients in DCSH.

1.2. Statement of problem

Tuberculosis (TB) disease remains a major public health problem which affects all age groups globally (FMOH, 2016). It is one of the top 10 causes of death and the leading cause from a single infectious agent and millions of people continue to fall sick with TB each year (Floyd et al., 2018)

The global TB report showed that there were an estimated 10 million incident cases and 12 million prevalent cases of TB. Overall 90% of the infection occurred in adults; of this nine percent were people living with HIV (72% in Africa). About 26% of incident TB cases occurred in Africa and 23% of the world's population are estimated to have a latent TB infection, and are thus at risk of developing active TB during lifetime (Floyd et al., 2018).

Currently, Ethiopia is ranked eighth among the 22 high TB burden countries in the world and at rank three, in Africa. The incidence rate of all forms of TB is estimated at 164 per 100,000 populations, leading to an annual mortality rate of 27.5 per 100,000 population (FMOH, 2016,Floyd et al., 2018).

A study which was conducted in Dessie city factors associated with treatment delay among newly diagnosed tuberculosis patients in Dessie city and surroundings, Northern Central Ethiopia by (Abdurrahman Seid and Yeshi Metaferia, 2018). using logistic regression model. since, the logistic regression model is not well suited to survival data for several reasons. According to (Collett, 2003), the survival times are not normally distributed and the censored data are the result of missing values on the dependent variable, but in this study the survival analysis method used to identify the risk factor of time to recovery.

Another study conducted by (Terefe and Gebrewold, 2018) using cox proportional hazard model survival analysis of time to recovery of tuberculosis patients in Mizan Tepi University Teaching Hospital, South-West Ethiopia, It depict the proportional hazard but not the survival time and the important covariate like residence ,WHO clinical stage ,marital status ,disease classification and age were not included. Therefore to fill this research gap the Accelerated failure time model

performs better than the proportional hazard model in applications where the effects of treatment are to accelerate or delay the event of interest (Kay and Kinnersley, 2002).

Therefore, this study addressed to answer the following research questions:

- What are factors that have significant effect on time to recovery of TB patients?
- What is the median recovery time of TB patients for the case of DCSH?

1.3. Objective of the Study

1.3.1. General objectives of study

The general objective of this study is to identify determinant factor of time to recovery of Tuberculosis patients under the follow up January 2017 to December 2020 at DCSH, South Wollo, Ethiopia.

1.3.2. Specific objective of the study

- ✓ To identify the factors influencing time to recovery of TB patients.
- ✓ To select parametric survival models that fit DCSH data.
- ✓ To estimate the median survival recovery time of Tuberculosis patients.

1.4. Significance of the Study

Tuberculosis treatment has the advantage of recovery the patients and indirectly preventing the transmission of the diseases from the patient to the normal persons. This study have great contribution for the patient and the TB professionals for follow up to eradicate the epidemic of the disease. It is also intended to forward doable recommendations to health institutions and policy makers on the way to increase recovery rate of TB patients. The study would be providing base-line data for detail and further studies in the future.

2. LITERATURE REVIEW

2.1. Definition and general overview of tuberculosis

Tuberculosis infection occurs when a person carries the tubercle bacilli inside the body but many people infected with tuberculous remain well for many years probable for life (Tizazu and Anteneh, 2006). In these asymptomatic individuals, the only evidence of infection is positive tuberculin skin test (WHO,2004).

2.1.1. Pulmonary Tuberculosis (PTB)

A patient with at least two sputum specimens which were positive for AFB by microscopy, or a patient with only one sputum specimen which was positive for AFB by microscopy, and chest radiographic abnormalities consistent with active pulmonary TB smear positive. A patient with symptoms suggestive of TB, with at least two sputum specimens which were negative for AFB by microscopy and with chest radiographic abnormalities consistent with active pulmonary TB or a patient with two sets of at least two sputum specimens taken at least two weeks apart, and which were negative for AFB by microscopy , and radiographic abnormalities consistent with pulmonary TB smear-negative and lack of clinical response to one week of broad spectrum antibiotic therapy (MOH,2008 ,Teklu, 1993).

2.1.2. Extra- pulmonary Tuberculosis (EPTB)

This included tuberculosis of organs other than the lungs, such as lymph nodes, abdomen, genitourinary tract, skin, joints and bones, meninges, etc. Diagnosis of EPTB was based on fine needle aspiration cytology or biochemical analyses of cerebrospinal/pleural/ascitic fluid or histo-pathological examination or strong clinical evidence consistent with active extra-pulmonary tuberculosis, followed by a decision of a clinician to treat with a full course of anti-tuberculosis chemotherapy. In all the cases of EPTB, sputum examinations and chest radiographs were used to investigate the involvement of lung parenchyma (MOH,2008 ,Belay et al., 2009).

2.1.3. Survival time

Tuberculosis is the major public health insult that attribute to 16 people per 100,000 populations. Mortality due to TB 1.7 death per 100,000 populations in WHO American region, which is the lowest of six WHO regions. The figure for European regional average is 3.7 per 100,000 populations (WHO,2015).

Africa highly affected by TB burden and contribute larger portion of illness and death to the world. WHO African region contain 12% of global population. However, contribute 31% of global TB burden. Moreover; WHO global TB report of 2015 for WHO African region reveal death of 46 individuals per 100,000 populations in the region, which is the highest for all regions (WHO,2015).

A retrospective cohort study conducted Among 1151 of TB patients registered during September 2011 to August 2016 and treated under the DOTS program using Kaplan Meier estimator and Cox PH regression models at MizanTepi University Teaching Hospital, South-West Ethiopia, Ethiopia (75%) were recovered and the rest (25.0%) censored from the study and the median recovery time of the patients was 185 days (6 months and 5 days) but it varies depending on the covariates included in the study. Among TB patients, 1139(99%) were new case patients and 12 (1.0%) were relapse. Among patients, 16 (1.4%) working place were from health facility and 1135 (98.6%) were not from health facility (Terefe and Gebrewold, 2018). People who are infected with HIV are 18 times more likely to develop active TB. The risk of active TB is also greater in persons suffering from other conditions that impair the immune system (WHO,2020).

A retrospective cross-sectional study conducted by (Assefa, 2019) on pulmonary tuberculosis and associated factors in Dessie referral hospital, Dessie, Ethiopia from September 1,2016 to august 31,2017 using binary logistic regression the prevalence of pulmonary tuberculosis in Dessie referral hospital was 67.5% relative to extra pulmonary tuberculosis the highest pulmonary tuberculosis patients was seen on the age group 31-45 and on sex group male. In addition to this the most pulmonary tuberculosis were HIV positive and lowest numbers of pulmonary TB were worked on health centre rather than non-health centre and most of TB was new case and there was

small number of relapse. weight, work area and HIV were significance associated with pulmonary TB occurrence.

2.2. Determinant factors for recovery time of tuberculosis

Majority of study findings claimed that death among TB patients while on treatment is not at random. Similar to survival time and rate of mortality, predictors of death among TB patients vary across six WHO regions in the world. Study conducted (Michael and Bolarinwa, 2020) on Survival Modelling of Tuberculosis Data-A Case Study of Federal Medical Centre, Bida, Nigeria by using Cox-proportional hazards model the result show that the older the TB patient, the lower was the time to recovery from TB. Males had higher hazards and hence, lower survival times compared to females. That is, female TB patients recovered faster than the males. The research concluded that age, gender and occupation were the major determinants of recovery period of TB patients.

Retrospective cohort study conducted by (Endris et al., 2014) on Treatment Outcome of Tuberculosis Patients at Enfraz Health Centre, Northwest Ethiopia, using logistic regression among 350 Among 417 study participants, 95 (22.8%), 141 (33.8%), and 181 (43.4%) were smear-positive, smear-negative, and extra pulmonary tuberculosis patients, respectively. there was no significantly association between sex, age, residence, type of TB, HIV status, and successful TB treatment outcome.

A study conducted by (Jakperik and Acquaye, 2013) on Assessing the Effects of Prognostic Factors in Recovery of Tuberculosis Patients in the Upper West Region from July 2007 to July 2012 the Product Limit function and Logistic Regression model result show that female patients had more chance of recovery than their male counterparts. Variables that significantly contributed to recovery were age, type of patents, duration of treatment, and HIV status of the patient with A median recovery time of TB was 25.43 weeks.

A retrospective cohort study conducted by (Geleso,2020) Based on the survival experience of TB patients, advancing age, extra pulmonary TB infection, living in rural residence, lower bodyweight

at beginning of treatment, HIV coinfection, and being a retreatment patient were predictors of mortality.

Study conducted by (Woldeamanuel and Mingude, 2018) on Factors Associated with Mortality in Tuberculosis Patients at Debre Berhan Referral Hospital, Ethiopia using logistic regression, the result shows the percentage of TB death was also higher among female participants and rural residents and there was no significance association between TB death and age, sex or residence ($p>0.05$). Studies done in Hawassa city showed that age, weight, smear negative pulmonary TB, dose of anti TB drugs, and HIV status were all factors associated with death of TB patients during the period of DOT (Debebe and Alemayehu, 2012).

Study conducted by (Balaky et al., 2019) on Survival analysis of patients with tuberculosis in Erbil, Iraqi Kurdistan region using cox regression model the result show that mean period of the follow-up of the patients was 7.6 months. Of 728 patients with tuberculosis, 50 (6.9%) had died. In multivariate analysis, patients with extra pulmonary disease (HR = 2.61, 95% CI 1.30–5.27).

Age: - As the age of a patient increases the chance of dying during TB treatment also increases (Biadglegne et al., 2013, Another study also as found from a study conducted in Dembia and Gonder which showed the death rate during TB treatment increased as age of patients increased (Tessema et al., 2011and Beza et al., 2013).similar findings result show that the higher hazards than younger ones, older patients had lower survival times. That is, the older the TB patient, the lower was the time to recovery from TB .The model suggested that recovery hazard of a male TB patient was 24.1% lower than that of a female patient, This implies that male patient had higher survival time than the female (Michael and Bolarinwa, 2020)

Body weight : finding show that the higher the body weight the faster the rate of the recover from TB (Terefe and Gebrewold, 2018)

Residence: Asebe et al., 2015 studied More TB patients were reported from urban than rural areas because of overcrowding, poverty and HIV infection. In the rural settings, access to the health service is limited; health seeking behaviour is poor and the living condition favour disease transmission. As a result, residential area indicating 98.61% from urban while 1.38% from rural area.

HIV co-infection: study conducted by (Beza et al., 2013) in Kolla Diba Health Centre result show that TB/HIV co-infected cases were more likely to die (13.3% vs. 2.0%) than HIV uninfected cases. TB-HIV co-infection has fatal consequences as TB becomes the leading cause of death in HIV infected individuals and patients with acquired immunodeficiency syndrome (AIDS). HIV lowers the host's immune response to MTB. The lifetime risk of developing active TB in HIV infected individuals is 10% per year compared with lifetime risk of 5 - 10% in individuals without HIV. As a result, the TB case notification rate (CNR) has increased four to six fold in sub-Saharan Africa (Shen et al., 2009, Deribew et al., 2009) . Another study result show that TB/HIV co-infected patients have much higher case fatality rate compared to HIV negative patients (Harries, 2009) . Another finding revealed that the recovery time of TB in HIV/TB co-infection was longer than that of HIV un-infected patients. In the absence of data from clinical trials, it is not known if duration of treatment of TB in HIV infected patients should be longer than in HIV un-infected patients. The few data that exist suggest that in HIV infected patient's duration of treatment for tuberculosis sensitive to first line therapy should be no difference to HIV uninfected patients BHIV, 2014.

A prospective cohort study in Spain reported that the risk of dying for TB/HIV co-infected patients was 7.08 times higher compared to HIV negative TB patients (Caylà et al., 2004). Similarly, another study also revealed that HIV infection had more than 5 times higher risk of mortality compared to HIV uninfected TB patients (Mugusi et al., 2009). Another study show that 86.1% were HIV positive and 13.9% were not (Terefe and Gebrewold, 2018).

2.2.1. Clinical (treatment) factors

Disease classification: study conducted by(Dioggban, et al., 2013) the Patients' diagnosis with extra pulmonary TB has more risk of treatment failure as compared to those with pulmonary TB. This is because it is very difficult to diagnosis the extra-pulmonary TB, patients have to go through series of laboratory test, ultra-scan and take X-rays to detect the affected area.

Treatment category: Among TB patients, 1139(99%) were new case patients and 12 (1.0%) were relapse (Terefe and Gebrewold, 2018). Another study patient who are newly diagnosis with TB have 98.5% chance of being cured than patients with relapse cases (Dioggban et al., 2013).

Type of TB: study in southern Ethiopia by (Datiko and Lindtjørn, 2010) type of TB is factors associated with mortality of patients who are on TB treatment. Being SNPTB case increased the risk of death compared to SPPTB and EPTB cases. But a similar study conducted in Addis Ababa presented with EPTB patients were more likely to die during TB treatment rather than SNPTB patients. The proportion of death from SPPTB, SNPTB and EPTB patients was 2.7%, 3.6%, and 4.3% respectively (Getahun et al., 2011).

Study conducted by (Belete et al., 2013) the mean of treatment success was not significantly affected by gender, age and type of TB. Based on the findings it is recommended to implement frequent supportive supervision during the course of treatment, strengthen referral linkage among facilities, and conduct further research to find out the reasons for the TB patients.

(Asebe et al., 2015) studied the Treatment outcome of Tuberculosis Patients in Gambella Hospital, Southwest Ethiopia by Using Bivariate analyses with a logistic regression model, the total of TB type 905 (78.29%) cases were PTB and 251 (21.71%) were EPTB patients from the study. Another study show that 49.04% of the patients had pulmonary tuberculosis whilst 50.96% had extra pulmonary tuberculosis (NTP, 2010; Jakperik and Acquaye, 2013. Another study show that patients infected with TB, 19.9% were extra pulmonary and 60.0%)were pulmonary. Extra pulmonary TB patients (21.1%) recovered the disease and Pulmonary TB patients 78.9% recovered the disease. Majority of patients with pulmonary TB recover the disease as compared to Extra Pulmonary TB patients (Terefe and Gebrewold, 2018)

Review of different literature as depicted in the above section, most of the studies listed in the literature used death as their event and the survival time determined based on the time to death of patients. While this study used recovery as the event and the survival time determined based on the time from date of admission to recover. This study therefore, is aimed at assessing the time to recovery among tuberculosis patients.

3. METHODOLOGY

3.1. Study area

The study was conducted in Dessie Comprehensive Specialized Hospital, north-eastern Ethiopia. Dessie city administration is found far from Bahir Dar (484 kilometers), which is the capital city of Amhara regional state,and 401 kilometers from Addis Ababa, the capital city of Ethiopia. It sits at a latitude and longitude of 11º8'N39º38E, with an elevation between 2470 and 2550 meters above sea level. The hospital has been given multiple services such as medical, surgical, obstetric, ANC, chronic follow-up, pediatric, orthopedic center service, DM service, ART and follow-up services, etc. According to the 2012 EC (Ethiopian Calendar) populations projection, there are 91870 households in Dessie. The total population of Dessie is 385850 (CSA,2019).

3.2. Study population and Source of data

All TB patients whose age above five years that have been admitted to DOT program from January 2017 to December 2020 at Dessie comprehensive specialised hospital were included in this study. The secondary data was directly extracted from the international WHO standard log Book which was adopted in 2008 to address the study variables. All the necessary data were collected from the follow up patients monitoring card and register book. Depending on the inclusion criteria stated below, all who fulfilled total of 311 Tuberculosis patients were included in the study. So, it was not much important for using further sample size determination methods. We simply take the total sample of 311 Tuberculosis patients which was recorded in the log-book.

Exclusion and Inclusion Criteria

The study those Patients who was referred to MDR-TB to other hospital and those who did not had complete history of their treatment in the log book were excluded from the study. The study had been included all TB patients whose age more than five years who had TB diseases record on the log book at Dessie comprehensive specialized hospital were included in the study.

3.3. Variables under the study

There are two types of variables for any statistical modelling which are known as dependent variable and the independent covariates.

Response Variable; The dependent variable of this study is time to recovery of tuberculosis patients, that is the length of time from the start date of taking drugs until the date of recovery. Patients who died, who are lost to follow up, dropped before recovery or transferred to other hospitals are considered as censored.

Predictor Variables; The predictor variables are assume to factor that affect recovery time of TB patients includes in the model are the following: -

Table 3.1 Predictor Variables

Table 3.1. 2 Demographic variables

Variables	Description of variables	Categories/value
Sex	Sex of TB patients	0=Male, 1=Female
Age	Age of TB patients	Continuous (year)
Residence	Residence of TB patients	0=rural,1=urban
Marital status	Marital status of TB patients	0=single,1=divorced /windowed,2=married

Table 3.1.3 Socio- economic variables

Variables	Description of variable	Categories/value
Initial weight	Initial weight of TB patients	Continuous(kilogram)
HIV status	HIV status of TB patients	0=HIV Negative, 1=HIV Positive
Work area	Work places of TB patients	0= health facility, 1 = Non-health facility

Table 3.1 4 clinical variables

Variables	Description of variable	Categories/value
Treat.cat	Treatment categories of TB patients	0 = New(N),1= relapse(R),
WHO clinical stage	WHO clinical stage of TB patients	0 = Stage I, 1 = Stage II, 2 = Stage III, 3= chronic
Disease classification	Disease classification of TB patients	0= pulmonary, 1= extra pulmonary
Types of TB	Type TB that exist of TB patients	0= Pulmonary Negative, 1= Pulmonary Positive, 2= extra pulmonary

3.4. Methods of data analysis

3.4.1. Survival analysis

The statistical method called survival analysis is appropriate to assess the stated objectives of this study. Survival analysis is used data in the form of times from a well-defined time origin until the occurrence of some particular event or end point.

A survival time is censored if all is known that is began or ended within some particular interval of time, and thus the total spell length (from admission time to recover) is not known exactly (Gardiner, 2010) .There are generally three reasons why censoring may occur:

A person does not experience the event before the study ends (right censoring), lost to follow-up during the study period (defaulters) and withdraws from the study because of death (if death is not the event of interest).

There are three categories of censoring

Right censored: The true observed event is the right of our censoring time. An observation is said to be right censored if it begins at time initial and terminate before the outcome of interest is observed (Elisa T. Lee & John Wenyu Wang, 2003) used to study response time data. In analysing such data, the main objects are to determine the length of time interval spent in a state and the transition probabilities from the current state to the entered state (Gharibvand and Liu, 2009).

Left censoring: Survival time is said to be left censored if an individual develops an event of interest prior to the beginning of the study.

Interval censoring: Survival time is said to be interval censored when it is only known that the event of interest occurs within an interval of time but the exact time of its occurrence is not known.

Survival Function

The survival function is defined as the probability that the survival time is greater or equal to t.

$S(t) = P(T \geq t), t \geq 0$ ----------(1)

$S(t)=1-F(t)$

Hazard Function

The hazard function gives the instantaneous failure rate at t given that the individual has survived up to time t, i.e.

$$h(t) = \lim_{\nabla t \to 0} P \frac{(t \leq T < \frac{\nabla t}{T} \geq t)}{\nabla t}, t \geq 0$$

$$h(t) = \frac{f(t)}{S(t)} = \frac{-dlogS(t)}{dt} \text{ ----------(2)}$$

3.4.2. Non parametric survival modelling

The Kaplan-Meier Estimator is a nonparametric estimator of the survival function which is not based on the actual observed event and censoring times, but rather on the order in which events occur will be applied to estimate the survival probability of the infected patients. This principle of nonparametric estimation of the survival function is to assign probability to and only to uncensored failure times. The log-rank tests are also applied to test the survival function by covariate values (groups).

The log rank test, developed by Mantel and Haenszel, is a non-parametric test for comparing two or more independent survival curves. It involves the calculation of observed and expected frequencies of failures in separate time intervals. Since it is a non-parametric test, no assumptions about the distributional form of the data need to be made. This test is however most powerful when used for non-overlapping survival curves. It can be generalized to accommodate other tests that are equally used sometime in practice such as Generalized Wilcoxon test, Tarone-Ware test, and

Peto-Peto Prentice test. Each of these tests uses different weight to adjust for censoring that is often encountered in survival data. The log rank test statistic for comparing two groups is given by:

$$X_{LR} = \frac{(\sum_{i=1}^{m} c_{1i} - \sum_{i=1}^{m} \hat{e}_{1i})^2}{\sum_{i=1}^{m} \hat{v}(\hat{e}_{1i})}$$

Where: m is the number of rank ordered event (recover) times, c_{ji} is the number of people experiencing the event at time $t_{(i)}$ in Group j, n_{ji} is the number of people at risk in group j at time $t_{(i)}$, c_i is the total number experiencing the event in both groups, $\hat{e}_{ji} = \frac{c_i n_{ji}}{n_i}$ is the estimated expected number of individuals experiencing the event at $t_{(i)}$ in group j,

$\hat{v}_{(\hat{e}_{ji})} = \frac{n_{1i} n_{2i} c_i (n_i - c_i)}{n_i^2 (n_i - 1)}$ is the estimated variance of $\hat{e}_{ji}$, n_i is the number of individuals at risk in both groups 1 and 2 just prior to event time $t_{(i)}$.

3.5. Survival Model

3.5.1. Semi-parametric methods

A Cox model is a statistical technique for exploring the relationship between the survival time and several explanatory variables. The most commonly used regression model is the Cox-proportional hazard model. With this model the distribution for the baseline hazard function is not specified implies vary with time and that is why it is called a semi-parametric model. The Cox-proportional hazard model is a more general model in modelling the hazard and survival function because it does not place distributional assumptions on the baseline hazard (Prentice,1992). The Cox model was introduced by Cox (1972) is written as:

$h_i(t \mid x) = h_o(t) \exp(X_i^t \beta)$ ---------- (3)

Where, $h_o(t)$ is the baseline hazard function; Xi is a vector of covariates and β is a vector of parameters for fixed effects. The corresponding survival function for Cox-PH model is given by:

$S(t,X) = [S_o(t)]^{\exp\{\sum_{i=1}^{p} \beta_i X_i\}}$----------(4)

where, $S_o(t)$ is the baseline survival function.

3.5.1.1. Methods of Parameter Estimation in cox Model

The regression coefficients in the proportional hazards Cox model, which are the unknown parameters in the model, can be estimated using the method of maximum likelihood. In Cox proportional hazards model we can estimate the vector of parameters β without having any assumptions about the baseline hazard $h_0(t)$.

Consider n independent individuals, the data that we need for the Cox proportional hazard model is represented by triplet (t_i, x_i, β_i), $i = 1, 2, \ldots n$ Where:

t_i is the survival time fori^{th}individual.

d_i is censoring for thei^{th}individual given by 0 for censored and 1 for event/death.

$x_i$isa vector of covariates for individual i^{th}then, the full maximum likelihood is defined as.

$$L(P) = \prod_{i=1}^{n} h(t_i, x_i, \beta_i)^{\delta_i} S(t_i, x_i, \beta_i) \text{---------- (5)}$$

Where

$h(t_i, x_i, \beta_i) = h_0(t)(e^{\beta' x_i})$ is the hazard function for individual i

$s(t_i, x_i, \beta_i) = S_0(t)(e^{\beta' x_i})$ is the survival function for individual i

When we rewrite the maximum likelihood it becomes

$$\boldsymbol{L(\beta)} = \prod_{i=1}^{n} \left(\left(h_0(t) \left((e^{\beta' x_i}) \right) \right)^{\delta_i} S_0(t)\, e^{\beta' x_i} \right) \text{---------- (6)}$$

Full maximum likelihood requires that we maximize with respect to the unknown parameter of interest, βunspecified baseline hazard and survival functions. This indicates that unless. We explicitly specify the baseline hazard; $h_0(t)$we cannot obtain the maximum likelihood estimators for the full likelihood. But, cox in 1972 proposed using an expression he called partial likelihood function that depends on only the parameter of interest.

Partial Likelihood

The general methodology used for proportional hazards, which cancels out the baseline function, is also used in determining the partial likelihood. To illustrate, the partial likelihood of an even to

occurring at time t for an individual we can be written as: P (individual I has experienced an event at time t (i) one event at time t (i))

$$L = \frac{h(t,xi,)}{\sum_{j=Rt(i)} h(t,xi,)} = \frac{\text{ho(t)}(e^{\beta' Xi})}{\sum_{j=Rt(i)} \text{ho(t)}(e^{\beta' Xi})} \text{---------- (7)}$$

When that there are no tied times assumed, the partial likelihood is defined over all failure time t (i) that i=1, 2, m & given as

$$\boldsymbol{L(\beta)} = \prod_{i=1}^{m} \frac{\beta' xi}{\sum_{j=Rt(i)}(e^{\beta' Xj})}$$

Where the product is over m distinct ordered failure times and $\boldsymbol{X}$ (i) denotes the value of the covariate for the subject with ordered survival time **t** (i). The log partial likelihood function is

$$\boldsymbol{L\ (\beta)} = \prod_{i=1}^{m} [\beta' xi - ln \sum_{j=Rt(i)}(e^{\beta' Xj})] \text{ ---------- (8)}$$

We obtain the maximum partial likelihood estimator by differentiating the right hand side of with respect to the component of β, setting the derivative equal to zero and solving for the unknown parameters, Breslow approximation used by different statistical software packages: in this study the Breslow approximation will be used.

The Breslow approximation

This approximation is proposed by Breslow and Peto to modify the partial likelihood and has the form

$$\boldsymbol{LB(\beta)} = \prod_{\boldsymbol{i=1}}^{\boldsymbol{m}} \frac{\boldsymbol{\beta' si}}{\sum_{\boldsymbol{j=Rt(i)}}(\boldsymbol{e^{\beta' Xj}})\mathbf{Ci}} \text{---------- (9)}$$

Where

- ✓ C the number of recovery occurred at time t_i
- ✓ $\boldsymbol{s}$ the sum of covariates over Ci subjects at time t_i

Then, the partial log of [10] is given as

$$\boldsymbol{LB(\beta)} = \prod_{i=1}^{m} [\beta' si - Ci \ln(\sum_{j=Rt(i)}(e^{\beta' Xj}))] \text{ ----------(10)}$$

Proportionality assumption of cox PH model

The basic assumption of Cox PH is proportional hazard which means the hazard ratio is constant over time. To check it the estimated survival curves cross, if they do, then this an evidence that the hazards are not proportional.

3.5.2. Accelerated Failure Time (AFT)Model

The Accelerated Failure Time model presents a way to easily describe and interpret survival regression data. It approaches the data differently than the widely used and well described Cox proportional hazard model, by assuming proportional effect of the covariates on the log-failure time rather than on the hazard function. It is an alternative if the proportional hazards assumption does not hold. Denote the survival functions of two groups by S_1 (t) and S_2(t), respectively, then the AFT model is given by:

$$S_1(t) = S_2(\phi t) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (11)$$

Where, $t \geq 0$ and ϕ is acceleration factor. This model implies that rate of group 1 is ϕ times as much as that of group 2. The hazard function of the i^{th} individuals at time t of the AFT regression model can be written in the form:

$$h_i(t) = e^{-\eta_i} h_o(t/e^{\eta_i}))$$

Where, $\eta_i = \alpha_1 x_{1i} + \alpha_2 x_{2i} +, \ldots, \alpha_p x_{pi}$ is the linear component of the model, in which x_{ji} is the value of the jth explanatory variable. X_j, j = 1, 2… p, for the ith individual, i = 1,2,…,n. The baseline hazard function $h_o(t)$ is the hazard of death at time t for an individual for whom the value of the p explanatory variables are all equal to zero.

The corresponding survivor function for the i^{th} individual is given by:

$$S_i(t) = S_o(t/e^{\eta_i})$$

Where, $S_o(t)$ is the baseline survival function. Let $\eta_i = \alpha' x_i$, then $e^{-\alpha' x_i}$ is the acceleration factor.

Notice that:

$e^{-\alpha' x_i} > 1$ implies that there is an acceleration of endpoint (recover).

$e^{-\alpha' x_i} < 1$ implies that there is a stretching or delay in endpoint(recover).

The log-linear model for a random variable T_i associated with the lifetime of the i[th] individual study, is given by;

$log_{T_i} = \mu + \alpha_1 x_{1i} + \alpha_2 x_{2i} +, \dots, + \alpha_p x_{pi} + \delta\epsilon_i$ ……..……..….(12)

The survival function of the random variable ϵ_i is given by:

$$S_i(t) = S_{\epsilon_i}\left(\frac{logt - \mu - \alpha_1 x_{1i} - \alpha_2 x_{2i} -, \dots, - \alpha_p x_{pi}}{\delta}\right.$$

Where, μ is intercept, δ is scale parameter and ϵ is the error distribution assumed to have a particular parametric distribution.

Similarly, the hazard function of the random variable ϵ_i is given by:

$h_i(t) = \frac{1}{\delta t} h_{\epsilon_i}\left(\frac{logt - \mu - \alpha_1 x_{1i} - \alpha_2 x_{2i} -, \dots, - \alpha_p x_{pi}}{\delta}\right)$

Where $h_{\epsilon_{i(\epsilon)}}$ is the hazard function of the distribution of ϵ_i.

Parametric accelerated failure time models based on Exponential, Weibull, Log-logistic and Log-normal distributions for survival time are the most commonly used.

3.5.2.1. Exponential Distribution

The exponential model reflects the property of the distribution appropriately called 'lack of memory. The parameter λ attains all positive values and the distribution with λ=1is called the unit or standard exponential. Therefore, the following formulae can be derived by some simple algebraic calculations:

Probability density function $f(x) = \lambda e^{-\lambda x}$,Survival function $S(x) = e^{-\lambda x}$

Hazard function$\lambda(x) = \lambda$, λ>0 , Cumulative hazard function Λ(x) = λx

$Mean\ E(X) = \frac{1}{\lambda}$, $Variance\ Var(X) = \frac{1}{\lambda^2}$.

The model is very sensitive to even a modest variation because it has only one adjustable parameter, the inverse of which is both mean and standard deviation.

3.5.2.2. Weibull Distribution

Weibull is a very flexible life distribution model. It has a hazard rate which is monotone increasing, decreasing, or constant. It is the only parametric regression model which has both a proportional hazards representation and an accelerated failure-time representation. The only difference between the Weibull model and the exponential model is that the scale parameter δ is estimated rather than being set to be one (Klein & Moeschberger, 2003). When T_i has a Weibull distribution then the survival function of the ϵ_i is given by:

$$S_{\epsilon_i}(\epsilon) = e^{(-e^{\epsilon})}$$

Where, $h_{\epsilon_i(\epsilon)} = e^{\epsilon}$. The survival function of the random variable T_i is given by:

$S_i(t) = e^{(-\lambda_i t^{\frac{1}{\delta}})}$ ----------(13)

With $\lambda_i = e^{(\frac{-\mu-\alpha_1 x_{1i}-\alpha_2 x_{2i}-,\ldots,-\alpha_p x_{pi}}{\delta})}$ and where λ_i is scale parameter, δ^{-1} is shape parameter. The hazard function of T_i is given by:

$h_i(t) = \frac{1}{\delta t}(\frac{logt-\mu-\alpha_1 x_{1i}-\alpha_2 x_{2i}-,\ldots,-\alpha_p x_{pi}}{\delta})$----------(14)

3.5.2.3. Log-Logistic Distribution

An alternative model to the Weibull distribution is the Log logistic distribution. The Log-logistic distribution has a fairly flexible functional form, it is one of the parametric survival time models in which the hazard rate may be decreasing, increasing, as well as hump-shaped that is it initially increases and then decreases (David Hosmer, et al., 2012)

When T_i has a Log-logistic distribution then the survival function of the ϵ_i is given by:

$S_{\epsilon_i(\epsilon)} = \frac{1}{1+e^{\epsilon}}$ ----------(15)

Where $h_{\epsilon_i(\epsilon)} = (1 + e^{-\epsilon})^{-1}$. The survival function for T_i is given by:

$S_i(t) = \{1 + e^{(\frac{logt-\mu-\alpha_1 x_{1i}-\alpha_2 x_{2i}-,\ldots,-\alpha_p x_{pi}}{\delta})^{-1}}$ ----------(16)

The hazard function of T_i for the i^{th} individuals is given by:

$h_i(t) = \frac{1}{\delta t}\{1 + [e^{-(\frac{\log t - \mu - \alpha_1 x_{1i} - \alpha_2 x_{2i} - ,\ldots, - \alpha p x_{pi}}{\delta})}]\}^{-1}$----------(17)

We can see that the log-logistic distribution has the proportional odds (PO) property. So this model is also a proportional odds model, in which the odds of an individual surviving beyond time t are expressed as

$\frac{S_T(t)}{1-S_T(t)} = \exp(-\alpha' x)\frac{S_o(t)}{1-S_o(t)}$----------(18)

3.5.2.4. Log-normal Distribution

When T_i has a Log-normal distribution then the survival function of the ϵ_i is given by:

$S_{\epsilon_i(\epsilon)} = 1 - \Phi(\epsilon)$……..…..…….……(19)

Where, $h_{\epsilon_i(\epsilon)} = \frac{e^{(-\epsilon^2/2)}}{(1-\Phi(\epsilon))\sqrt{2\pi}}$. The survival function of T_i for the i^{th} individuals is given by:

$S_i(t) = 1 - \Phi(\frac{\log t - \eta_i - \mu}{\delta})$ ……..…….(20)

Where, with parameters $\eta_i - \mu$ and δ.

3.6. AFT parameter estimation

Accelerated failure time models are fitted using maximum likelihood estimation. The likelihood of the n observed survival time $t_1, t_2, \ldots, t_n$ is given by:

$L(\alpha, \mu, \delta) = \prod_{i=1}^{n}[f_i(t_i)]^{\delta_i}[S_i(t_i)]^{1-\delta_i}$----------(21)

The log likelihood function is then:

$$\log L(\alpha, \mu, \delta) = \sum_{i=1}^{n}\{-\delta_i \log(\delta t_i) + \delta_i \log f_{\epsilon_i}(z_i) + (1-\delta_i)\log S_{\epsilon_i}(z_i))\}$$

Where, $z_i = \frac{\log t - \mu - \alpha_1 x_{1i} - \alpha_2 x_{2i} - ,\ldots, - \alpha_p x_{pi}}{\delta}$, $S_i(t_i) = S_{\epsilon_i}(z_i)$.

The maximum likelihood is manipulated by Newton-Raphson procedures using software.

3.7 Variables selection in the study

According to Hosmer and Lemeshow (1998), it is recommend to follow the steps in selecting the variables by:.

1. The first step is to fit model that contain each of the variables one at a time.

2. We begin by fitting a multivariable model containing all variables significant in the univariable analysis at the 20-25 percent level, as well as any other variables not selected with this criterion but judged to be of clinical importance.

3. Use backward selection to eliminate non-significant variables and examine the effect of remaining variables.

4. Fit the final model by omitting variables that are non-significant and adding variables that are significant.

3.8 Model selection

3.8.1. Akaike's information criterion (AIC)

For comparing models that are not nested, the Akaike information criterion (AIC) can be used (Akaike, 1974) and it is defined as:- AIC = -2(LL) + 2(k + c) ---------- (22)

Where LL is log-likelihood, k is the number of covariates in the model and c is number of shape parameters in model.

3.8.2. Bayesian Information Criteria (BIC)

The Bayesian Information Criteria (BIC) is given by Schwarz (Schwarz, 1978). It is computed as follows

$$BIC = -2(log - likelihood) + (P + K) * \log(n)$$----------(23)

Where P is the number of parameters in the distribution, K is the number of coefficients and log(n) is the number of observations. The distribution which has the lowest BIC value is considered as best fitted model.

3.9. Model diagnosis

Model Diagnostics for survival data the evaluation of model adequacy are often based on quantities known as residuals. Residuals for survival data are slightly different than for other types of regression models, due to censoring.

3.9.1. Assessing the proportional hazards assumption

The assumptions of Cox-PH model was checked by test of correlation (rho) and global test. The assumption is valid using test correlation (rho) and global test if the test result is insignificant. The graphical methods can be used to check if a parametric distribution fits the observed data.

3.9.2. Model checking in parametric models

Datwyler and Stucki (2011) represents to use the plots;

- log[S(t)] versus t gives approximately a straight line that pass through the origin if exponential distribution is reasonable.
- log[−log[S(t)]] versus log[t] gives approximately a straight line (linear) if Weibull distribution is reasonable.
- log[S(t)/1 − S(t)] versus log(t) is linear if log-logistic distribution is reasonable.
- ϕ^{-1} [1−S(t)] versus log(t) is linear if log-normal distribution is reasonable.

Where

- S(t) denotes the probability that a randomly selected individual survives from the time origin to sometime t or beyond time t.
- $\phi(.)$ denotes the standard normal distribution function.

3.9.3. Cox Snell Residual

The residual that is most widely used in the analysis of survival data is the Cox-Snell residual, so called because it is a particular example of the general definition of residuals given by Cox and Snell (Cox & Sell 1968; Cox & Oakes, 1984). The Cox-Snell residual for the i^{th} individual, i =

1,2, … , , is given by properties and features of residuals, when survival outcome are modelled, have been extensively studied in the literature.

S(t:X)= $[(S_o t)]^{\exp(\beta\chi)}$ Or, in terms of hazards: h (t; X)= $h_o(t)\exp(\beta\chi)$. so for each person with covariates x_i S(t:x_i) = $[(S_o t)]^{\exp(\beta x_i)}$ then we can calculate $\widehat{h_\iota}$ = −log$[\hat{S}(T_i; x_i)]$ Or first predict survival probability at the actual survival time for individual, then log-transform it. The residuals in right censored data constitute a censored sample of the unit exponential distribution $r_{ci} = \widehat{H_\iota}(ti^*) = -log\widehat{S_\iota}(ti^*)$

Where $\widehat{H_\iota}(ti^*)$ and $\widehat{S_\iota}(ti^*)$ are the estimated cumulative hazard and survivor functions respectively, for the i^{th} individual at the censored survival time. then the model cox-Snell residual is given by

rci'= 1-δi+rci ----------(24)

3.10. Operational definition of terms

When patients are entered to DOTs they categorized to one of the treatment category. According to the national TB/Leprosy control program guideline (FMOH, 2013). this treatment categories are:-

New case: A patient who never had treatment for TB, or has been on TB treatment for less than four weeks.

Treatment after failure: A patient who, while on treatment, is smear positive at the end of the fifth month or later, after commencing.

Transfer in: A patient who is transferred-in to continue treatment in a given treatment unit after starting treatment in another treatment unit.

3.11. Ethical considerations

Permission to undertake this study is obtained from College of Natural Science through Ethical Review Board and official letter of co-operation is written by Debre Berhan university research Directorate to Dessie Comprehensive Specialized hospital. After the data is collected, Confidentiality of the information and privacy of the center (both the staff and cases from card) is secured.

4. RESULT AND DISCUSSION

4.1 Descriptive results

Out of 311 registered TB patients 130(41.80%) were recovered and 181(58.20%) were censored during the follow up period. 186 (59.8%) patients were females and 125(40.19%)were males. These patients had a median, mean and standard deviation survival recovery of 9,10.54, 5.01, months respectively. Of 311 TB patients registered for the DOT program 163(52.41%) of the patients were urban area while 148(47.59%) were rural areas. The percentage of patients in WHO clinical stage I, II, III, and IV were 53 (40.77) ,27 (20.77), 36 (27.69) and 14 (10.77) respectively. About 170 (54.66%) of the patients had Pulmonary TB, while 141 (45.34%) were diagnosed with extra-Pulmonary TB. From the total of 311, 141(45.34%) of the TB patients were single while 121 (38.91%) of the TB patients were married, and the remaining 49(15.76%) were divorced/windowed.

From the total of 311, 95(30.55%) of the patients had HIV positive, while 216(69.45%) were had HIV negative. 55(42.31%) of rural residents were recovered from TB during the study while 75(57.69%) of the patients were recovered from the urban patients. The recovery for marital status of single, married and divorced/windowed were 68(52.31%),37(28.46%) and 25(19.23%) respectively. 84 (64.62%) of the patient's health facility were recovered, while 46 (35.38) were non health facility recovered, the recovery for type of TB were 41(31.54%),31 (23.85%) and 58(44.62%) of pulmonary negative, pulmonary positive and extra pulmonary respectively.

Table 4. 1: Time to recovery TB patients in DCSH from 2017-2020GC

Demography and health factor		total (%)	recovery/ event(%)	Censored(%)
Sex	Female	186(59.81)	76(58.46)	110(60.77)
	Male	125(40.19)	54(41.54)	71(39.23)
Residence	Rural	148(47.59)	55(42.31)	93 (51.38)
	Urban	163(52.41)	75(57.69)	88 (48.62)
Marital status	Single	141(45.34)	68(52.31)	73 (40.33)
	Married	121(38.91)	37(28.46)	84(46.41)

	Divorced	49 (15.76)	25(19.23)	24 (13.26)
disease classification	Pulmonary	170(54.66)	70 (53.85)	100 (55.25)
	extra pulmonary	141(45.34)	60 (46.15)	81 (44.75)
type of TB	Pulmonary negative	82(26.37)	41 (31.54)	41(22.65)
	pulmonary positive	88(28.30)	31 (23.85)	57 (31.49)
	extra pulmonary	141(45.34)	58(44.62)	83 (45.86)
treat. category	new case	250(80.39)	104(80.66)	146 (80.00)
	Relapse	61 (19.61)	26 (19.34)	35 (20.00)
work area	non health facility	105(33.76)	46 (35.38)	59 (32.60)
	health facility	206(66.24)	84 (64.62)	122 (67.40)
WHO clinical stage	stage I	119(38.26)	53 (40.77)	66(36.46)
	stage II	88(28.30)	27 (20.77)	61 (33.70)
	Stage III	69 (22.19)	36(27.69)	18.23 (33)
	Chronic	18(11.25)	14(10.77)	21(11.60)
HIV status	HIV positive	95 (30.55)	48 (36.15)	47 (26.52)
	HIV negative	216(69.45)	83(63.85)	133 (73.48)

As show in below table, the median period TB patients were 9 months. The minimum follows up time was 1 month and the maximum time was 36 months. The average Age and Weight of the patients' in the study were 33 years and 44kgs respectively. The Minimum Age of patients was 5 years and the maximum Age of patient was 89 years. The minimum body Weights of patients were 6.5kgs and maximum body Weights of the patients were 90kgs respectively.

Table 4 .2 The descriptive statistics of continuous covariates

variable	Total	mean	S. deviation	Min.	Max.	median
weight	311	44.3828	17.53075	6.5	90	45
Age	311	33.24116	16.39521	5	89	30
time (in months)	311	10.54341	5.008755	1	36	9

4.2 The Kaplan-Meier Estimator

Survival time distributions of time-to-recovery of TB patients were estimated for each group using KM method. From figure 4.1 below of KM curve, the horizontal axis shows that the time-to-recovery from tuberculosis, whereas the vertical axis shows the probability of survival $p(T > t)$. At the beginning, the survival curve is increasing, at t= 0 S(t) = 1 implies that when the patient are at medium stages they have high probability of recovery whereas at the end the survival curve is decreasing, and implies that the probability of recovery of TB decreases when time increases.

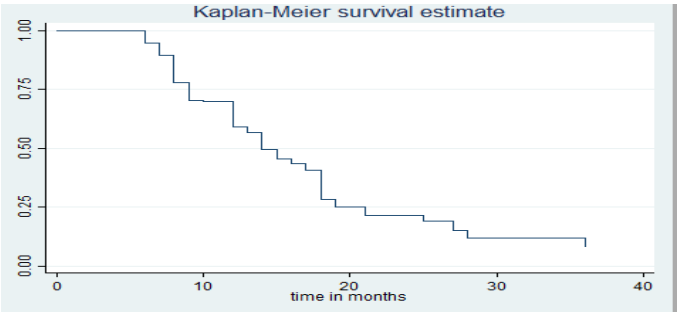


Figure4. 1 Estimated of survival function of TB patients

Individual graph of kaplan meier :-

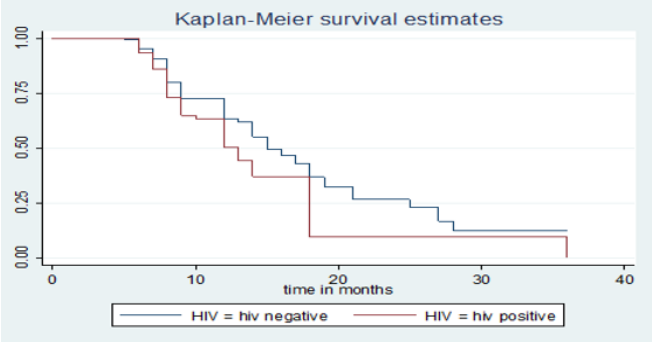


Figure 4.2 the plot of the estimate of Kaplan-Meier survivor function of recovery time of tuberculosis patients in DCSH by HIV status

From the figure 4.2, the HIV negative is more accelerated to recover than HIV positive. we observed that the two survivor curves were the same until the 5^{th} months when the curve for the HIV positive started a gradual fall. The curves are distinguishable and serve as a confirmation of the earlier result that there was statistically significant difference in recovery times between HIV

negative and HIV positive patients of TB. However, the other variables like sex, WHO clinical stage, treatment categories and work area see "Appendix 3 A".

Log rank test for survival curves of TB patients

The log-rank test was used at 5% level of significance to validate the differences in the time to recovery of each factor. There is no difference between the probabilities of an event occurring at any time point was the null hypothesis that has been tested. The corresponding log rank test of equality of the survival curves of covariates is summarized below.

Table 4.3 The log-rank test among the categorical variables of TB data in DCSH

Variables	DF	Chi-square	p-value
Sex	1	1.3091	0.076
Treatment Category	1	22.2618	0.502
Residence	1	28.005	<0.032
marital status	2	20.321	0.06
Disease classification	1	36.087	< 0.04
type of TB	2	38.405	<0.00
WHO clinical stage	3	29.6486	<0.00
HIV status	1	15.431	<0.00
work area	1	2.346	<0.01

From table of the log-rank test type of TB, HIV status, disease classification, residence, work area and WHO clinical stage have p -value less than alpha value 0.05. This implies that there is difference between the probabilities of an event occurring at any time point of the study time. while sex, treatment category and marital status were not statistically significantly different. Controlling for sex, the log rank test statistic (1.3091, Pr=0.076) showed that there was no significant difference in recovery times for both sexes and controlling for marital status, the log rank test statistics (20.321, pr=0.06) show that there was no significance difference in recovery times for all categories of marital status.

4.3 Inferential statistics

4.3.1 Cox Proportional Hazard Model

4.3.1.1. Univariable Analysis

Univariable analysis used to realise effect of covariates on recovery time of tuberculosis patients and to select variables to be included in the multivariable analysis. The result of univariable analysis indicated that age, residence, work area, type of TB, WHO clinical stage, disease classification, HIV status and weight were significant for Cox proportional hazards model at 25% level of significance while treatment category, marital status and sex were not significance ("Appendix 1 A"). Therefore, all significance predictor variables were included in the multivariable analysis

4.3.1.2. Multivariable Analysis

The multivariable analysis of recovery time of Tuberculosis patients using Cox proportional hazards model was fitted by including all significant covariates in the univariable analysis. Using the backward elimination method to select the final significant covariates age, disease classification, type of TB, HIV status, work area, WHO clinical stage and weight were significant at 5% level of significance in "Appendix 1B".

4.3.1.3. Assumption checking for Cox proportional hazard model

While the Goodness of fit testing approach is employed for Cox PH model, The PH assumption was checked by using graphical method and tests based on the Schoenfeld residual.

Test of proportional hazard assumption by schoenfeld residual

The Schoenfeld residual is one of the methods used to check the PH assumption. In this study the p-value was checked for testing the assumption is fulfilled or not. The global test result in ("Appendix 1C") shows that p-value is significant (p-value=0.0004). This implied that there is evidence to contradict the proportionality assumption. So the proportionality assumption is not fulfilled.

Test of proportional hazard assumption by graphical method

the graphs for covariates are not parallel, implying that the proportional hazards assumption not satisfied. This also supports the above result that is the PH assumption is violated for this TB dataset ("Appendix 3B"). Thus, the assumption of Cox PH model violate and AFT model consider for this data set.

4.3.2 Accelerated Failure Time Models

The univariable analyses were fitted for each covariate with a p-value less than 0.25 by using different AFT models such as Exponential, Weibull, Log-logistic and Log-normal distribution. The AFT models show that the covariates like age, residence, weight, disease classification, work area, type of TB, HIV status, and WHO clinical stage are found to be significant with time to recovery of TB patients at 25% level of significance.

For time to recovery of TB data, Multivariable analysis of AFT models were fitted using all significant covariates in univariable analysis. Using the backward elimination method to select the final significant covariates. The covariates such as age, disease classification, type of TB, HIV status, work area, WHO clinical stage and weight were significant in all AFT models at 5% level of significance. The model comparison was done using those significant covariates for each AFT models.

4.3.2.1 Model Selection

From table 4.4 the value of AIC and BIC for all AFT models are displayed. the smallest AIC and BIC show that the 'best' fit model for data analysis. Then log logistics with accelerated time has smallest AIC value and BIC value. In addition to AIC, the standard error of log logistics AFT model was smaller standard error than the other base line AFT models ("Appendix2 B"). Thus indicates the log-logistic AFT model better to fits the TB patients' data. Therefore, the analysis of this study was performed by using log logistics regression model with AFT.

Table 4.4 Comparisons of AFT models using statistical information criteria

information criteria	Common base line AFT models			
	Exponential	Weibull	Log normal	**Log logistics**
AIC	504.8937	374.6167	351.105	**342.2307**
BIC	546.0314	419.4942	395.9825	**387.1082**

4.3.2.2 Log-logistic Accelerated Failure Time Model

From table 4.5 below result of multivariable log-logistic AFT model is presented. The accelerated factor greater than one ($\phi>1$) indicates prolong time to recover while the accelerated factor ($\phi<1$), indicates that they have shortened time to recover as compared to their reference category. Work area, HIV status, disease classification and WHO clinical stage (stage II and III) significantly shorten time-to-recover, but chronic not significantly different than other stages. The estimated acceleration factor of HIV status of TB patients was 0.8586 than the reference (HIV negative). Which indicates that the HIV positive patients decreased by a factor of 0.8586. The estimated acceleration factor for type of TB was 1.115 of pulmonary positive using pulmonary negative as a reference category. This indicates that for pulmonary positive were increased by a factor 1.115 than the reference group (pulmonary negative) when other covariates constant. When the age of the patients increases by one year, the acceleration factor (ϕ) 1.017 times prolong time-to-recovery. Based on log logistic model, an acceleration factor (ϕ) of patients extra pulmonary was 1.129 times prolong time to recovery than the reference group (pulmonary) meaning that extra pulmonary have 12.9% higher length of time to recovery than pulmonary TB (reference). When initial weight of the patients increased by one unit, the acceleration factor (ϕ) 1.055 times prolong time-to-recovery.

Table 4. 5 Result of multi variable analysis of Log Logistic AFT model

Covariates	Categories	ϕ	SE	p-value	95%Conf. interval	
Type of TB	Pulmonary negative	Ref				
	Pulmonary Positive	1.1153	.0103	0.000	1.0517	1.1721
	Extra Pulmonary	1.1323	.0561	0.025	1.1152	1.1946
HIV status	HIV negative	Ref				

	HIV positive	.8586	.0633	0.001	.7430	.9921
Work area	health facility	Ref				
	Non Health facility	.8866	.0629	0.001	.7715	.9189
Declassification	Pulmonary	Ref				
	Extra Pulmonary	.1295	.0772	0.025	.1164	.2814
Age		1.0171	.0024	0.000	1.0113	1.0257
Weight		1.0556	.0022	0.000	1.0341	1.0742
WHO	stage I	Ref				
clinical	stage II	.9682	.0877	0.003	.8106	.9968
stage	stage III	.7776	.0759	0.001	.6421	.9417
	chronic	.7894	.1007	0.475	.6146	1.0138
Intercept		14.7152	0.9676	0.000	11.3242	19.1216

95% CI =confidence interval for acceleration factor; SE: standard error. Ref: reference

4.4. Model Diagnostics

After fitting the model, we should check the goodness of fit of the model whether it fits the data well or not. We use both the formal (global test of BETA=0) likelihood ratio test and the most common graphical method, (cox Snell) of checking the goodness of fit of the final selected model.

4.4.1. The Cox Snell Residual Plots

As shown in the figure 4.5 bellow, The Cox Snell residual plot, From the graph we can observe that the cumulative hazard function follows the 45^0 line except for large values of time. Therefore, we can conclude that the final log-logistics AFT model fits the data very well.

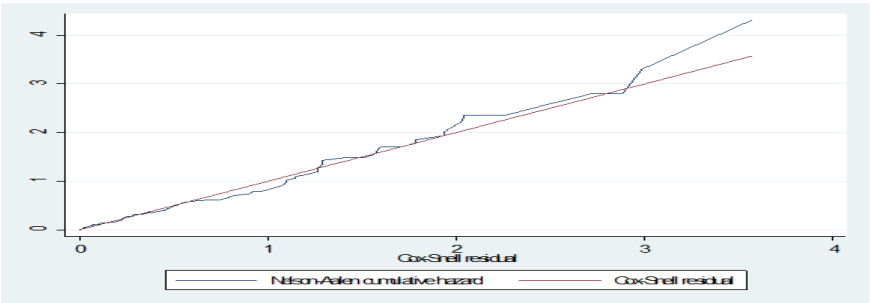


Figure 4. 3 Cumulative hazard plot of the Cox-Snell residual for log-logistic AFT model

4.4.2. Likelihood Ratio Test for log-logistic AFT Model

From table 4.6, the likelihood ratio test at 10 degree of freedom is 22.05 and the p- value is 0.0148, the likelihood value of the full model and null model were -159.1153 and -170.1419 respectively. Thus, imply that the model fits the data very well.

Table 4. 6 The likelihood ratio test for the log-logistic AFT model

Loglike (model)	Loglike (intercept only)	χ^2	df	p-value
-159.1153	-170.1419	22.05	10	0.0148

4.5. DISCUSSION

The aim of the study was to identify determinant factors of time to recovery of Tuberculosis patients under the follow up January 2017 to December 2020 at DCSH, South Wollo, Ethiopia. The Cox-PH model was first used for this data. But the assumption of proportionality in Cox-PH model was violated, the AFT models with common baseline distribution: Exponential, Weibull, Log-logistic and Log-normal were considered. To compare different AFT models AIC and BIC was used and log-logistic AFT model is found to be the 'best' fit for the time to recovery of TB patient's data than others. The multivariable analysis using the log-logistic analysis showed age, disease classification, work area, type of TB, WHO clinical stage, HIV Status and weight were significantly affect the time to recovery of TB infected patients.

The estimated acceleration factors of HIV status having HIV positive is 0.86 with 95%CI [0.7430 0.9921] which indicates that HIV positive have shorter survival time to recovery than that of HIV negative (reference category). Which is in line with Caylà et al., 2004 and Mugusi et al., 2009. Finding in Kola Diba Health Centre by Beza et al., 2013 result show that TB/HIV co-infected cases were more likely to die (13.3% vs. 2.0%) than HIV uninfected cases. TB-HIV co-infection has fatal consequences as TB becomes the leading cause of death in HIV infected individuals and

patients with acquired immunodeficiency syndrome (AIDS). People who are infected with HIV are 18 times more likely to develop active TB WHO,2020. Which is consistent with this finding. Since, death and recovery are controversial.

The finding of the study shows that 41.80% were recovered and the rest 58.20% were censored and the median time to recovery of the patients was 270 days (nine months). which is contradict with study in South-West Ethiopia by Terefe and Gebrewold, 2018 the result revealed that (75%) were recovered and the rest (25.0%) censored from the study. Their difference may be due to different pattern of TB distribution with different study duration.

The finding of the study Patients diagnosed with extra pulmonary TB decreased by 0.12 times for recovery time as compared to those with pulmonary TB holding the effect of all covariates constant. The study result shows that, the Extra-pulmonary TB needs shorten time to recovery than pulmonary TB WHO, 2013. It needs an experience doctor to confirm and at times involve the use of an X-rays imaging or scans and tuberculin skin test among others Konstantinos, 2010. Another study conducted in Addis Ababa presented with EPTB patients were more likely to die during TB treatment rather than SNPTB patients. The proportion of death from SPPTB, SNPTB and EPTB patients was 2.7%, 3.6%, and 4.3% respectively Getahun et al., 2011.which is consistent with this finding.

The finding show that 105 (33.76%) work area were non health facility and 206(66.24%) were health facility. Which is contradict with the study done in Terefe and Gebrewold, 2018. The reason of the study finding may be due to different pattern of TB distribution with different places. The acceleration factor of work area is 0.8886 meaning that non health facility have 88% decreased length of time to recovery from TB patients than the health facility (reference).

The finding of study the acceleration factor of age is 1.017, this indicates the age of the patient increase the time to recovery from TB is increase. Which is contradict to Debebe and Alemayehu, 2012, Michael and Bolarinwa, 2020, Tessema et al., 2011 and Beza et al., 2013 The difference of the study may be the mean age of this study is 33 years. This indicates that they have the disease at an early age, so they often have no difficulty to recovery. Most papers reported that TB attacks adults than children because of the more contacts and it mode of transmission.

In this study shows that 55% of the respondents had pulmonary TB while 45% of the respondents had extra pulmonary tuberculosis. The result is in line with Assefa,2019, Balaky et al.,2019 and Endris et al., 2014 that proves 19.9% were extra pulmonary and 60.0% were pulmonary and Extra pulmonary TB patients 21.1% recovered the disease and Pulmonary TB patients 78.9% recovered the disease.

In this study initial weight increased by one unit had a 1.055 times higher rate of recovery. This finding is consistent with a study from Mizan Tapi University by Terefe and Gebrewold, 2018 revealed that higher the body weight the faster the rate of the recover from tuberculosis.

5 CONCLUSION AND RECOMMENDATION

5.1. CONCLUSION

From the result of AIC and BIC the log logistic AFT model provided a suitable choice in order to model time of recovery from TB data obtained from Dessie comprehensive Specialized Hospital as compared to other base line AFT models. The median time to recovery from TB was 270 days which was approximately nine months. Age of patent, initial weight, HIV status, WHO clinical stage, disease classification, work area and types of TB are significant factors for time recovery from tuberculosis. The Kaplan-Meier method was used to estimate the time to recovery of patients after beginning of TB treatment. The Log Rank test shown that initial weight, disease classification, type of TB, WHO clinical stage of TB, residence, age, HIV status and work area had significantly contributed effect on recovery time of TB patients. However, sex of the patients, marital status and treatment categories of the patient have insignificant difference for recovery time of tuberculosis patients.

Recovery rate of pulmonary positive patients prolong time to recovery than pulmonary negative patients. Similarly, the recovery rate of extra pulmonary patients prolong time to recovery than the reference (pulmonary negative). The recovery rate of non-health facility patients were shorten time to recovery than health facility of TB patients (reference). The recovery rate of patients under WHO clinical stage (II and III) is short time to recover as WHO clinical stage 1 reference.

5.2. RECOMMENDATION

Based on the finding of the study, we recommended as follows; -

- The policy makers and health service providers should give special treatment for HIV positive patients, since they shorten time to recovery.
- For governmental and non-governmental organization responsible should be give attention for health facility in order to reduce death timing of TB patients and enough martials to treatment.

REFERENCE

ADDO, K., YEBOAH-MANU, D., DAN-DZIDE, M., OWUSU-DARKO, K., CAULLEY, P., MENSAH, G., MINAMIKAWA, M., LIENHARDT, C., BONSU, F. & OFORI-ADJEI, D. 2010. Diagnosis of tuberculosis in Ghana: the role of laboratory training. Ghana Medical Journal, 44.

AKAIKE, H. 1974. A new look at the statistical model identification. IEEE transactions on automatic control, 19**,** 716-723.

ASEBE, G., DISSASA, H., TEKLU, T., GEBREEGIZEABHE, G., TAFESE, K. & AMENI, G. 2015. Treatment outcome of tuberculosis patients at Gambella Hospital, Southwest Ethiopia: three-year retrospective study. Journal of Infectious Diseases & Therapy.

ASSEFA, E. 2019. Pulmonary Tuberculosis and Associated Factors in Dessie-referral Hospital, Dessie, Ethiopia. *J Health Med Informat*, 10**,** 2.

BALABANOVA, Y., RADIULYTE, B., DAVIDAVICIENE, E., HOOPER, R., IGNATYEVA, O., NIKOLAYEVSKYY, V. & DROBNIEWSKI, F. A. 2011. Survival of drug resistant tuberculosis patients in Lithuania: retrospective national cohort study. BMJ open, 1.

BATES, M. N., KHALAKDINA, A., PAI, M., CHANG, L., LESSA, F. & SMITH, K. R. 2007. Risk of tuberculosis from exposure to tobacco smoke: a systematic review and meta-analysis. Archives of internal medicine, 167**,** 335-342.

Belay Tessema, Abebe Muche, Assegedech Bekele, Dieter Reissig, Frank Emmrich, and Ulrich Sack, (2009).Treatment outcome of tuberculosis patients at Gondar UniversityTeaching Hospital, Northwest Ethiopia. A five - year retrospective stud: BMC Public Health 9:371

BERHE, G., ENQUSELASSIE, F. & ASEFFA, A. 2012. Treatment outcome of smear-positive pulmonary tuberculosis patients in Tigray Region, Northern Ethiopia. *BMC public health*, 12**,** 1-9.

BEZA, M. G., WUBIE, M. T., TEFERI, M. D., GETAHUN, Y. S., BOGALE, S. M. & TEFERA, S. B. 2013. A five years tuberculosis treatment outcome at Kolla Diba Health Center, Dembia District, Northwest Ethiopia: a retrospective crosssectional analysis.

BHIV, (2014). British Health Report on TB/HIV. WWW. BHIVA - 4.0 Type and duration of TB treatment.htm

BIRUK, M., YIMAM, B., ABRHA, H., BIRUK, S. & AMDIE, F. Z. 2016. Treatment outcomes of tuberculosis and associated factors in an Ethiopian University Hospital. *Advances in Public Health*, 2016.

COLLETT, D. 2015. Modelling survival data in medical research, CRC press.

Collett, David, (2003). Modeling Survival Data in Medical Research. Chapman and Hall, London.

CSA (Ethiopia). Population projections for Ethiopia - central statistical agency, 2019. Available: http://www.csa.gov.et/census-report/

COX, D. R. 1972. Regression models and life-tables. Journal of the Royal Statistical Society: Series B (Methodological), 34**,** 187-202.

Cox, (1972). Regression models and life tables with discussion Journal of the Royal Statistical Society. Cox, D.R. and Oakes, D., (1984). Analysis of survival data. London: Chapman and Hall.

CDC. (2014). Centers for Disease Control and Prevention. Treatment of Tuberculosis, American Thoracic Society, CDC, and Infectious Diseases Society of America.

DATIKO, D. & LINDTJØRN, B. 2010. Mortality in successfully treated tuberculosis patients in southern Ethiopia: retrospective follow-up study. The international journal of tuberculosis and lung disease, 14**,** 866-871.

DERIBEW, A., TESFAYE, M., HAILMICHAEL, Y., NEGUSSU, N., DABA, S., WOGI, A., BELACHEW, T., APERS, L. & COLEBUNDERS, R. 2009. Tuberculosis and HIV co-infection: its impact on quality of life. Health and quality of life outcomes, 7**,** 1-7.

DINGETA, T. & ABDOSH, T. 13. Risk Factors for Unsuccessful Tuberculosis Treatment Outcome (Failure, Default and Death) in Selected Public Health Institutions, Eastern Ethiopia, 2012. Office of the Vice President for Research Affairs**,** 106.

DIOGGBAN, J. 2011. SURVIVAL ANALYSIS OF AVERAGE RECOVERY TIME OF TUBERCULOSIS PATIENTS IN NORTHERN REGION, GHANA.

Dioggban.j and Michael o., (2012). Survival analysis of average recovery time of tuberculosis patients in northern region, Ghana, International journal of current research.

David W. Hosmer, Stanley Lemshow, Susanne May, (2012). Survival analysis: Regression modeling of time to event data. 2nd ed. Wiley series.

DR, C. & OAKES, D. 1984. Analysis of survival data.

Elisa T. Lee & John Wenyu Wang, (2003). Statistical method for survival data analysis, 3rd ed, John wiley inc

Elvan Akturk Hayat, Asli Suner, Burak Uyar, Omer Dursun, Mehmet N. Orman and Gul Kitapcioglu MD., (2010). Comparison of Five Survival Models: Breast Cancer Registry Data from Ege University Cancer Research Center; Turkiy, J Med Sci, 30(5), 1665-74.

ENDRIS, M., MOGES, F., BELYHUN, Y., WOLDEHANA, E., ESMAEL, A. & UNAKAL, C. 2014. Treatment outcome of tuberculosis patients at Enfraz Health Center, Northwest Ethiopia: a five-year retrospective study. Tuberculosis research and treatment, 2014.

FAURHOLT-JEPSEN, D., RANGE, N., PRAYGOD, G., JEREMIAH, K., FAURHOLT-JEPSEN, M., AABYE, M. G., CHANGALUCHA, J., CHRISTENSEN, D. L., GREWAL, H. M. & MARTINUSSEN, T. 2013. Diabetes is a strong predictor of mortality during tuberculosis treatment: a prospective cohort study among tuberculosis patients from M wanza, T anzania. Tropical Medicine & International Health, 18**,** 822-829.

FLOYD, K., GLAZIOU, P., ZUMLA, A. & RAVIGLIONE, M. 2018. The global tuberculosis epidemic and progress in care, prevention, and research: an overview in year 3 of the End TB era. The Lancet Respiratory Medicine, 6**,** 299-314.

FOK, A., NUMATA, Y., SCHULZER, M. & FITZGERALD, M. 2008. Risk factors for clustering of tuberculosis cases: a systematic review of population-based molecular epidemiology studies. The International Journal of Tuberculosis and Lung Disease, 12**,** 480-492.

GARDINER, J. Survival analysis: overview of parametric, nonparametric and semiparametric approaches and new developments. SAS Global Forum 2010. Statistics and Data Analysis, 2010. 1-23.

Gavrilenko, V. S., (2001). Recovery criteria and time in patients with pulmonary tuberculosis. Tuberk, (8): 10-4.

GEBREZGABIHER, G., ROMHA, G., EJETA, E., ASEBE, G., ZEMENE, E. & AMENI, G. 2016. Treatment outcome of tuberculosis patients under directly observed treatment short course and factors affecting outcome in southern Ethiopia: a five-year retrospective study. *PloS one,* 11**,** e0150560.

GETAHUN, B., AMENI, G., MEDHIN, G. & BIADGILIGN, S. 2013. Treatment outcome of tuberculosis patients under directly observed treatment in Addis Ababa, Ethiopia. *Brazilian Journal of Infectious Diseases,* 17**,** 521-528.

GEBREEGZIABHER, S. B., YIMER, S. A. & BJUNE, G. A. 2016. Tuberculosis case notification and treatment outcomes in West Gojjam Zone, Northwest Ethiopia: a five-year retrospective study. *Journal of Tuberculosis Research,* 4**,** 23-33.

GELESO, M. G. 2020. Modeling the Survival of Tuberculosis Patients in Eastern Zone of Tigray Regional State. *Risk Management and Healthcare Policy,* 13**,** 473.

HARRIES, A. D. 2009. Reducing case Fatality in HIV-infected TB Patients in sub-Saharan Africa. Southern African Journal of Epidemiology and Infection, 24**,** 10-12.

HEALTH, F. D. R. O. E. M. O. 2013. Guidelines for clinical and programmatic management of TB, TB/HIV and Leprosy in Ethiopia. Federal Democratic Republic of Ethiopia, Ministry of Health Addis Ababa

HEALTH, F. M. O. 2016. Guidelines for clinical and programmatic management of TB,

TB/HIV and Leprosy in Ethiopia. FMoH Addis Ababa, Ethiopia.

JAKPERIK, D. & ACQUAYE, B. K. 2013. Assessing the effects of prognostic factors in recovery of tuberculosis patients in the upper west region. Mathematical Theory and Modeling, 3.

JONNALAGADA, S., HARRIES, A. D., ZACHARIAH, R., SATYANARAYANA, S., TETALI, S., CHANDER, G. K., RAO, S., RAO, R., PERI, S. & ANCHALA, R. 2011. The timing of death

in patients with tuberculosis who die during anti-tuberculosis treatment in Andhra Pradesh, South India. BMC public health, 11**,** 1-7.

Kalbfleisch, J. D., & Prentice, R. L. (2011). The statistical analysis of failure time data (Vol. 360). John Wiley & Sons

Kaplan, E. L., and P. Meier. (1958). Nonparametric estimation from incomplete observations. Journal of the American Statistical Association 53: 457–481.

Klein, J. P., & Moeschberger, M. L. (2003). Survival analysis: techniques for censored and truncated data. Springer Science & Business Media.

Klembaum, D. G., (1996). Survival Analysis: A Self learning text. Springer, New York

KONSTANTIN 2010. spread of Tubercluosis.

KAY, R. & KINNERSLEY, N. 2002. On the use of the accelerated failure time model as an alternative to the proportional hazards model in the treatment of time to event data: a case study in influenza. Drug information journal, 36**,** 571-579.

LIN, H. & ZELTERMAN, D. 2002. Modeling survival data: Extending the cox model. Taylor & Francis.

Lawless, J. F., (1982). Statistical Methods and Model for Lifetime Data. Wiley, New York.

LÖNNROTH, K., JARAMILLO, E., WILLIAMS, B. G., DYE, C. & RAVIGLIONE, M. 2009. Drivers of tuberculosis epidemics: the role of risk factors and social determinants. Social science & medicine, 68**,** 2240-2246.

MCSHANE, H. 2005. Co-infection with HIV and TB: double trouble. International journal of STD & AIDS, 16**,** 95-101.

MICHAEL, V. A. & BOLARINWA, I. A. 2020. Parametric Survival Modeling of Tuberculosis Data-A Case Study of Federal Medical Centre, Bida, Nigeria. Modern Applied Science, 14.

MILLER 1983. methods of parametric model estimation.

Ministry of Health of Ethiopia (MOH), (2008). Tuberculosis, Leprosy and TB/HIV Prevention and Control Programme Manual. Addis Ababa: MOH 4th edition.

MUNIYANDI, M., RAMACHANDRAN, R., GOPI, P., CHANDRASEKARAN, V., SUBRAMANI, R., SADACHARAM, K., KUMARAN, P., SANTHA, T., WARES, F. & NARAYANAN, P. 2007. The prevalence of tuberculosis in different economic strata: a community survey from South India. The International Journal of Tuberculosis and Lung Disease, 11**,** 1042-1045.

National Tuberculosis Control Programme (NTP). 2010. Annual Tuberculosis Report, Ghana

ORGANIZATION, W. H. 2013. Global tuberculosis report 2013, World Health Organization.

ORGANIZATION, W. H. 2018. Technical report on critical concentrations for drug susceptibility testing of medicines used in the treatment of drug-resistant tuberculosis. World Health Organization.

PAN, W. 2001. Akaike's information criterion in generalized estimating equations. Biometrics, 57**,** 120-125.

PARDESHI, G. 2009. Survival analysis and risk factors for death in tuberculosis patients on directly observed treatment-short course. Indian journal of medical sciences, 63.

REVES, R. & ANGELO, S. 2016. As Ethiopia moves toward tuberculosis elimination, success requires higher investment, JSTOR.

SAMUEL, L., HIKO, D. & GIRMA, A. 2016. Survival status and its predictors among patients with tuberculosis in hosanna, southern Ethiopia: retrospective cohort study.

SILVA, D. R., MENEGOTTO, D. M., SCHULZ, L. F., GAZZANA, M. B. & DALCIN, P. T. 2010. Mortality among patients with tuberculosis requiring intensive care: a retrospective cohort study. BMC Infectious Diseases, 10**,** 1-7.

TEKLU, B. 1993. Symptoms of pulmonary tuberculosis in consecutive smear-positive cases treated in Ethiopia. Tubercle and Lung Disease, 74**,** 126-128.

TEREFE, A. & GEBREWOLD, L. 2018. Modeling Time to Recovery of Adult Tuberculosis (Tb) Patients in Mizan-Tepi University Teaching Hospital, South-West Ethiopia. Mycobact Dis, 8**,** 2161-1068.1000258.

TESSEMA, B., MUCHE, A., BEKELE, A., REISSIG, D., EMMRICH, F. & SACK, U. 2009. Treatment outcome of tuberculosis patients at Gondar University Teaching Hospital, Northwest Ethiopia. A five-year retrospective study. *BMC public Health*, 9**,** 1-8.

THACKER, S. B., QUALTERS, J. R., LEE, L. M., CONTROL, C. F. D. & PREVENTION 2012. Public health surveillance in the United States: evolution and challenges. MMWR Surveill Summ, 61**,** 3-9.

Therneau, T.M and P.M. Grambsch, (2000). Modeling Survival Data: Extending the Cox Model. Springer Verlag, New York.

WANG, J. & SHEN, H. 2009. Review of cigarette smoking and tuberculosis in China: intervention is needed for smoking cessation among tuberculosis patients. BMC public health, 9**,** 1-9.

WOLDEAMANUEL, G. & MINGUDE, A. 2018. Factors associated with mortality in Tuberculosis patients at Debrebirhan referral hospital, Ethiopia: A Retrospective Study. J Trop Dis, 7**,** 2.

World Health Organization, (2020). Global tuberculosis report. WHO 2006 report.

XUE, X., XIE, X., GUNTER, M., ROHAN, T. E., WASSERTHEIL-SMOLLER, S., HO, G. Y., CIRILLO, D., YU, H. & STRICKLER, H. D. 2013. Testing the proportional hazards assumption in case-cohort analysis. BMC medical research methodology, 13**,** 1-10.

APPENDICES

Appendix 1

A) univariable analysis of cox proportional model at 25% level of significance

Covariate	categories	Haz.Ratio	Std. Err.	P-value	[75%Confi.interval]

Sex	female	Ref.				
	male	1.0362	.1855	0.842	.8433	1.2733
Residence	rural	Ref.				
	Urban	1.0224	.2214	0.010	1.0087	1.0613
Marital status	single	Ref				
	Marriage	.7667	.1855	0.249	.5804	1.0128
	Divorced	.9368	.3037	0.598	.6451	1.3603
Age		.8946	.00821	0.003	.4352	.9604
HIV	HIV negative	Ref.				
	HIV positive	1.5197	.3026	0.025	1.3285	1.7454
Work area	health F.	Ref.				
	Non h facility	1.3224	0.2359	0.101	1.077059	1.62375
Weight		.4199	.0063	0.000	.3926	.5272
WHO clinical stage	stage I	Ref.				
	Stage II	1.1373	.3040	0.002	1.0359	1.1665
	StageIII	2.0979	.5710	0.007	1.9334	2.1703
	Chronic	1.7223	.4163	0.021	1.1411	1.8995
Type of TB	pu. negative	Ref.				
	Pulmonary positive	. 7658	.1696	0.03	.5935	.9881
	extra Pulmonary	.7158	.1803	0.104	.5357	.9564
Treat. cate.	new cases	Ref.				
	Relapse	1.4428	.5626	0.347	.9212	2.2596
Declass	Pulmonary	Ref				
	Extra Pulmonary	1.5755	.2006	0.001	1.3905	1 .6499

Ref; reference

B) multivariable cox proportional hazard result at 5% level of significance

Covariates	Categories	Haz. Ratio	Std. Err.	P-value	95% Conf. Interval	
Age		.9128	.0081	0.000	.8570	.9885

HIV status	HIV negative	Ref				
	HIV positive	1.5591	.3098	0.005	1.2405	1.9597
Weight		.6791	.0063	0.004	.5679	.9857
WHO clinical Stage	Stage I	Ref				
	Stage II	1.0519	.2681	0.030	1.0389	1.1237
	Stage III	2.0096	.5229	0.041	1.0067	2.0467
	Chronic	1.3721	.6430	0.05	.5476	1.4378
Type of TB	Pul. negative	Ref				
	Pul.pulmonary	.8091	.1817	0.010	.6709	.9565
	Extra. Positive	.7265	.1810	0.004	.5452	.8154
Disease classification	Pulmonary	Ref				
	E. pulmonary	1.5892	0.2066	0.002	1.3536	1.8820
Work area	Health Facility	Ref				
	Non Health facility	1.3558	0.2599	0.025	1.2904	1.6933

C) Test of proportional-hazards assumption by sechonefeld residual

Covariate	Rho	chi-square	P-value
Age	-0.1415	2.55	0.1100
Weight	-0.07335	0.63	0.4280
HIV	-0.00462	0.15	0.9572
Type of TB	0.00836	0.01	0.9189
WHO clinical stage	-0.02739	0.10	0.005
Disease classification	0.03446	0.15	0.6970
Work area	0.09724	1.27	0.2602
global test	NA	8.43	0.0004

Appendix 2

A) Univariable log logistics AFT Model at 25% level of significance

Covariate	Categories	Coef.	Std. Err.	P-value	[75% Conf.Interval]	
Sex	Female	Ref.				
	Male	.017	.069	0.803	-.097	.062
Residence	Rural	Ref.				
	Urban	-.015	.069	0.122	-.094	-.005
Marital status	Single	Ref.				
	Marriage	.085	.077	0.272	-.039	.175
	Divorced	-.025	.095	0.790	-.135	.084
Age		.009	.002	0.001	.001	.033
HIV	HIV negative	Ref.				
	HIV positive	-.140	.072	0.052	-.223	-.056
Work area	Health facility	Ref.				
	Non health facility	-.077	.071	0.077	-.160	-.048
Weight		0.023	.0019	0.006	.014	.049
WHO clinical	Stage I	Ref.				
Stage	Stage II	-.016	.087	0.052	-.017	-.014
	Stage III	-.200	.096	0.039	-.311	-.086
	Chronic	-.244	.122	0.045	-.385	-.104
Type of TB	Pulmonary negative	Ref.				
	Pulmonary positive	.071	.080	0.137	.021	.163

	Extra pulmonary	.107	.088	0.125	.005	.208
Disease Classifications	pulmonary	Ref.				
	Extra pulmonary	.163	.075	0.030	.076	.250
Treatment Categories	New cases	Ref.				
	Relapse	0.382	.085	0.401	-.020	.216

Ref; reference

B) Multivariable AFT model at 5% level of significance

		Distributions			
		Exponential	Weibull	Log-normal	Log-logistic
Covariate	categories	$\hat{\beta}$ [S.E]	$\hat{\beta}$ [S.E]	$\hat{\beta}$ [S.E]	$\hat{\beta}$ [S.E]
Age		.015 (.0026)	. 012 (.0028)	.011 (.0035)	.0169 (0.0024)
HIV status	HIV negative	Ref			
	HIV .pos.	- .3291 (.1894)	-.1794 (.0797)	-.1594 (.0734)	-.1524(.0633)
Work area	health facility	Ref			
	Non health Facility	-.2799 (.1805)	-.1920 (.0742)	-.1255 (.0713)	-.1203(.0629)
Weight		.0201(.0065)	.0251 (.0059)	.015 (.0071)	.0541 (.0022)
WHO clinical stage	Stage I	Ref			
	Stage II	-.3712(.2464)	-.0472(.1058)	-.0565(.0956)	-.0323(.0877)
	Stage III	-.7227 (.2637)	-.3279(.1120)	-.2780 (.1031)	-.2515(.0759)
	Chronic	-.5714 (.3487)	-.2690 (.1432)	-.2663 (.1301)	-.2264(.1007)
Type of TB	Pulmonary negative	Ref			
	Pulmonary Positive	.1044 (.0969)	.0808(.1863)	.1060 (.0827)	.1091 (.0103)

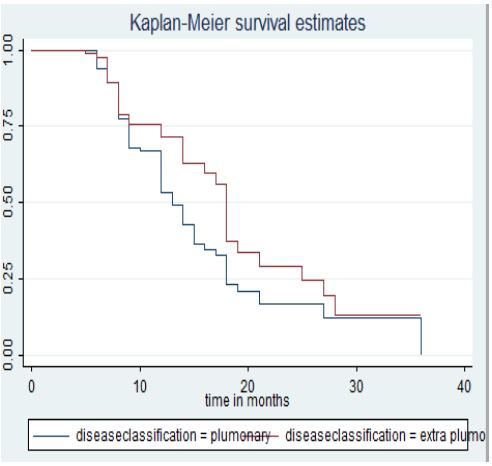
	Extra pulmonary	.1092 (.2368)	.1089(0.2141)	.1109 (.0924)	.1242 (.0561)
Disease classification	Pulmonary	Ref			
	Extra pulmonary	-3.1293(.2041)	-2.1104(.0830)	-3.1121(.1083)	-2.0440 (.0772)

Ref; reference

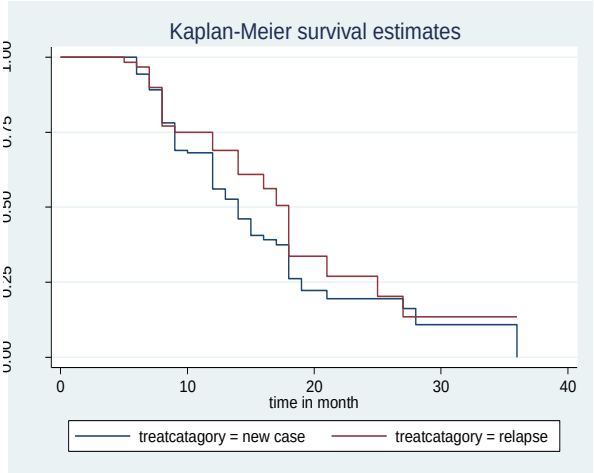
Appendix 3

A**) the plot of the estimate of Kaplan-Meier survivor function with different covariates**.

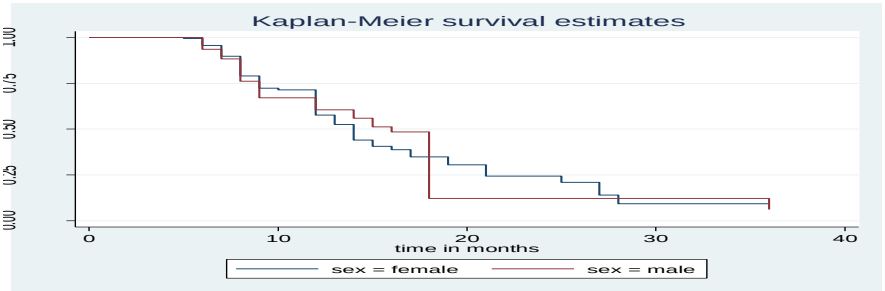
plot of WHO clinical stage



plot of treatment category

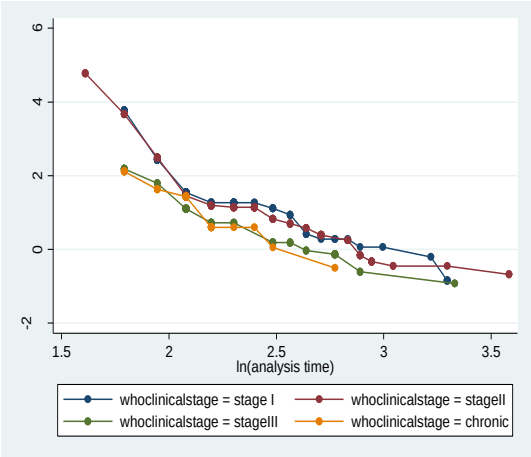


plot of sex



B) PH assumption by graphical

by WHO clinical stage



by Type of TB